SPACE TIME ANALYSIS IN AN ENTERPRISE GIS

Ritesh Agrawal, University of Illinois Urbana-Champaign
Email: ragrawal@uiuc.edu

ABSTRACT:

The spatial and non-spatial data is increasingly used in the disease ecology and ecosystem health analysis and management. The present state of spatial data management system and the currently available tools are restrictive in developing and analyzing complex models. Recognizing the need for better and efficient data management system and specialized tools, the College of Veterinary Medicine at University of Illinois, Urbana-Champaign initiated a project to build a central GIS data repository and develop specialized temporal analysis tools. This paper discusses the development of central spatial data repository and the temporal analysis tools that will be used by the College of Veterinary Medicine. This project is unique since it is the first initiative towards building the central database repository for the college research projects.

INTRODCUTION:

With the increasing importance of spatial data in ecosystem analysis and its relation to health issues, enterprise geographic information systems (GIS) offer a new way of sharing and accessing spatial and non-spatial data across groups and institutions. With ever expanding options, the integration of geographic information systems and database management systems is now possible.

In the last few decades there has been an increased research interest in disease ecology and ecosystem health analysis and management. Research in disease ecology and ecosystem health uses spatial data to examine and analyze patterns and trends to develop complex models from a wide array of parameters, including various aspects of soils, geology, and topography as well as human settlements and infrastructure. The temporal and spatial components are often both important in such analyses, yet the temporal dimension of spatial data has often been under developed. Yet, since these parameters are subject to changing over to time and are highly related to a variety of disease processes and ecosystem health concerns, temporal analysis is one of the most important aspects of the research. At the same time, the temporal dimension of spatial data is under developed (1).

Temporal analysis requires storing data for a parameter over different time period, contributing to an immense amount of spatial data. The maintenance of such massive amount of spatial data is a mammoth task and requires meticulous planning. Responding to the urgency of managing the spatial and temporal data, the College of Veterinary Medicine, at the University of Illinois Urbana Champaign embarked on a project titled "Illinois Ecosystem Health Mapping System". The objective of the project is to build a central spatial data repository for research project carried out in the Veterinary College using enterprise geographic information system (GIS). In addition, specialized tools are being developed to aid better visualization of temporal and spatial changes and the development of risk maps.

Enterprise GIS offers a new way of sharing and accessing spatial and non-spatial data over groups and institutions (3). While many enterprise systems are developed for counties, cities or other

governmental units, many research groups fail to organize their own data this way. The objective of the project is to build central data repository and spatial temporal analysis tools according to users needs and requirements. The direct benefactors of this project are the students and faculty staff of the Veterinary College.

METHODOLOGY:

The project is being carried out in 4 overlapping sequential phases. The four phases are:

1. Development of goals and objectives.
2. Assessment of user needs.
3. Design system architecture.
4. Implementation of the system.

Phase 1. Development of goals and objectives:
The objective of the project, "Illinois Ecosystem Health Mapping System" is to create a centralized GIS data repository conforming to the requirements of the research conducted at the College of Veterinary Medicine in the areas of ecosystem health and disease ecology. The most benefit was gained in a centralized system that focused on the state of Illinois, where a considerable amount of research has been carried out on such issues as amphibian decline, the mapping and analysis of West Nile virus and other human and animal diseases.

The objectives of the project beside managing spatial and non-spatial data are:

1. ***Reduce Processing Time:*** Locating, acquiring and formatting the spatial data is the most costly and long-drawn-out part of any ecosystem health management project. Preparing the spatial data in a most coherent format is a cumbersome process and causes unnecessary delay in the research projects. The delays to the projects are mainly caused because of the following reasons (6):
   - Finding appropriate source for the data
   - Incomplete/Insufficient information about the data
   - Data transfer problem
   - Projection adjustment

   By maintaining a data repository all these problems are solved. Since all the available spatial and temporal data is obtained, updated, processed and projected into a single projection system. So the personnel in the college can get spatial data in a ready-to-use format.

2. ***Promote re-use of spatial data*****:** By storing the spatial and non-spatial data in the central data repository the problem of duplication of data can be solved. Two or more projects requiring the same data can very efficiently share the data. In addition, field works that produces new data can be loaded to the system and shared across groups and projects.

3. ***Provide specialized tools and user interface*****:** The need for and experience with spatial data and GIS software varies widely among the users of the college. Some of the personnel do not use GIS software while for other users using spatial data is a small part of their project. A small group of power users provide intensive processing and spatial analysis. Based on project requirements and personal experience it was clear that different users have different expectation with the data, data search and tools for spatial temporal analysis. According to the needs of the

veterinary research projects it was decided to develop some customized tools to speed up the temporal analysis process and to facilitate searching the central data repository.

Phase 2. Assessments of user needs:
The end users for this data repository are the personnel related to Veterinary Medicine. Based on the experience and knowledge in using the spatial data, need for the spatial and non spatial data in the project, and the level of expertise in GIS software, the users are categorized into three groups:

1. ***GIS Analysts***: They specialize in using GIS software and do not require any specific tool or user interface to access or analyze data. At the same time, repetitive tasks that are common to GIS data processing may be reduced through well-designed supplemental tools. They have the highest expectation for speed and efficiency from an enterprise system and are often the ones that people come to when they need help finding or processing spatial data.
2. ***General Users:*** They have the ability to perform any mapping but do not specialize in GIS. Any specialized tool or interface to access and analyze data will certainly help them and increase their speed but it is not mandatory to have a specialized tool or interface.
3. ***Occasional Users:*** These users occasionally access data for their own work. They may or may not be specialized in GIS. The Specialized User Interface will be most helpful for the occasional user.

To evaluate the varying needs of the different types of users, we have led discussions regarding the type of data of most interest to the most users. Based on initial discussions, the following decisions were taken

- ***Spatial Extent***: The focus of the major research work going in the veterinary medicine is concerned with the state of Illinois in a broad sense and in particular with Chicago area and areas adjacent to Illinois River basin. To increase the data search efficiency and also of the system architecture, it is decided to store the spatial data in three different spatial extents:
  - The Chicago area
  - The Illinois River Basin
  - The state of Illinois
- ***Data format:*** The projection of the final data is an important decision as it affects the final map products. The projection used widely throughout state government is the Lambert Conformal Conic projection with custom parameters for the state. The data will be in North American datum of 1983. Therefore it is decided to re-project all the data into this projection system before storing the spatial data in the database.

Phase 3. System Architecture:
The recent software and hardware advances have provided better network access to spatial data, allowed more complex geospatial data models, and have made the integration of GIS and database management systems a reality (4). System architecture design provides a hardware and network solution that supports the performance and communication needs of the users of GIS. Well-planned system architecture is indispensable for an efficient enterprise GIS. Well-planned system architecture significantly improves the transmission of data across the organization. After evaluating the available options within the budget constraints, it was decided to use "Oracle 9i" database management system along with ESRI's ArcSDE product as the middleware for connections to the spatial data. Both Oracle and ArcSDE reside on a Red Hat Linux data server.

The ESRI software is not yet fully supported under Linux, but through a number of trials and attempts, both ArcSDE and the ESRI Internet Map Server (ArcIMS) were successfully installed under Linux. This platform forms the framework for a variety of projects that are carried out in the state of Illinois. While web mapping was not a part of the enterprise system described here, it is anticipated that connections to the data repository through ArcIMS will also be of use to users in the future.

Phase 4. Implementation of the System

The implementation of the system is a multifold task requiring proper documentation at each step. Some of the crucial steps are:

Data Inventory and Data Search: The data repository was built by consolidating data from a variety of sources and projecting all the data in one projection system. Only those data which included metadata were included in the system. In ArcSDE, all data with the same spatial extent are stored in a separate database. To keep the speed of retrieval high, each of the three regions of interest to the EHMS was stored in a separate database. The three are all within one ArcSDE service. This solution has a weakness in that the same data are stored more than one time at the different levels of geography. While not ideal, this is expedient and since computer storage space was not an issue, it was carried forward.

For the study area in Chicago, special care was taken to create a biologically significant area that included both natural and urban area (Figure 1). The initial data for this area are Ikonos panchromatic and multi-spectral images. These images provide a base for field work and a means to better map vegetation and natural features in the area. It is anticipated that future field work and more detailed data for this area will be added as students use this region for field activities related to wildlife and disease ecology in this region.

Data for the Illinois River basin are available through the Army Corps of Engineers Restoration Needs Assessment (RNA) project (http://www.mvr.usace.army.mil/ILRiverEco/default.htm). This project resulted in a set of data for this area that was fairly comprehensive from the point of view of ecology. The Illinois River basin includes numerous biologically significant areas and has been an important focus of river restoration activities. Because the area, or a portion of it, is frequently put forward as a study area for ecosystem projects, it is included in the enterprise system. One challenge in using the data from the RNA was that data for the entire state are replicated in the RNA data. Only one set of metadata were deemed needed for each specific layer, however, to avoid giving the user the impression that the soils (or other) data with the Illinois River basin region was different from that at the state level.

A key aspect of the final implementation is improved tools for data discovery. Storing the data in the central data repository is of no use if the user cannot know whether the data they are looking for is available or not. To overcome the difficulty of finding a data, it was decided to store information of all the spatial data in an excel sheet. The parameters to be stored in the excel sheet were conformed to Dublin Core metadata standard. The Dublin Core standard comprises fifteen elements, the semantics of which have been established through consensus by an international, cross-disciplinary group of professionals from librarianship, computer science, text encoding, the museum community, and other related fields of scholarship (9). Also each spatial data is classified into a broad category and then into a subclass. Appendix A Table 1 shows the classification that was used.

A Database Builder was written to update the excel sheet automatically. This Database Builder automatically takes the required parameters from the metadata and updates the sheet while the other parameters like theme, Rights, original formats, contributors etc are given by an administrator. The routine is written by using ArcObjects and Visual Basic and is used in ArcCatalog.

User Interface: For preliminary search and viewing the files in SDE database an application "EHMS Viewer" is being developed. This application apart from giving common options like Zoom in, Zoom out, identity, etc gives the option to search the SDE database on the basis of category and subclass as described above. Though all the functions except the search engine are available in all GIS software, this independent system is provided for those who do not have GIS software on their computer or who need to determine what is in the repository. Also this application is more specialized according to needs of the users and therefore the occasional user of GIS can easily adapt to this application as compared to other GIS software. Appendix A Fig 4 shows the EHMS Viewer application. An additional feature of this application is that it allows the user to query data by a date field and do initial exploration of temporal trends.

For disease ecology and ecosystem health analysis and management, temporal analysis is very important. The present GIS software doesn't provide any option for temporal analysis and if they do the process is very cumbersome. To overcome this difficulty, a tool "AniMapper" has been developed using ArcObjects and Visual Basic within ArcGIS 8x. The tool is being developed within the ArcGIS environment so that the user can utilize the other wide variety of function already available in ArcGIS. This tool has functions to query data for a variety of temporal periods based on a date field. For example, in one database, data on West Nile virus comes from a period of several months. The cases may be viewed by day, week, month, year or a custom, user-defined period.

After selecting the Starting and ending data and the increment period, the user can see the selected records after each increment of time by clicking on next or previous buttons. Also at any moment he can export the selected records and shape it as a shape file. This greatly reduces the time needed to create an animated series of maps. Apart from visual display the tool also creates an excel sheet and a graph displaying the number of records found per time interval and in cumulative time interval. The advance options allow the user to select some particular field on which he want to do temporal analysis. Appendix A Fig. 3 shows the AniMapper, the tool developed by GISSA lab for temporal analysis.

CONCLUSION

The Illinois Ecosystem Health Mapping System is the first initiative towards building the central database repository for the college research projects. Apart from building the central data repository, tools have been developed for temporal analysis, for entering metadata, and for searching and viewing spatial data. Building a single database repository is a complex project in light of the lack of standards, variability in the requirements of the user, and technology limitation. The various decisions taken while building database repository are affected by the available funds, end-users and technology available. Since these parameters change from agency to agency, the same standards and decisions might not be applicable to other institutions. This project is unique as it focuses on research requirements in an academic environment, it integrates environmental and socio-demographic data with human and animal health data, and it provides solutions for visualization and access to both temporal and spatial data. The future objective of this project lies in developing a set

of tools that will automate the frequently occurring tasks related to analysis and visualization of space-time changes, including preparing data for use in the specialized statistical packages that can test for spatial and temporal clusters

References:

1. Langran, Gail. 1992. *Time in Geographic Information System*. London: Taylor & Francis

2. ESRI White Paper. 1999. Enterprise GIS in Health and Social Service Agency

3. Von Meyer, Nancy R. and R. Scott Oppman, ed. 1999. *Enterprise GIS*. Park Ridge, IL: URISA.

4. Ruiz, M.O, *et al.*, 2001. Data Enterprise Repository Prototype for Fort Hood, Texas. ERDC Technical Report, May 2001.

5. Giacinto, J.F., Cutler, W.G., Hrovatic, M.D. 2003. Development, Implementation, and Maintenance of Geographical Information System with Remote Query Capabilities. Available online at URL:http://gis.esri.com/library/userconf/proc00/professional/papers/PAP176/p176.htm

6. Cote, P. 1999. Fundamentals of GIS Lecture Outline. Available online at URL:http://www.gsd.harvard.edu/~gsd6322

7. Ku, W., Wu, J. The application of GIS for Infectious Disease. Available online at URL:http://gis.esri.com/library/userconf/proc00/professionals/papers/PAP283/p283.htm

8. Pearson, Chris. 2003. Towards a Single Zone Projection for Illinois. Available online at URL: http://www100.state.il.us/ilgic/projection.cfm

9. Ibaugh, Allen. 1999. Interdepartmental GIS Data Sharing in Osceola County, Florida: The Utilization of Internet/Intranet Data Visualization Tools. Available online at URL:http://www.asu.edu/caed/proceedings99/IBAUGH/IBAUGH.HTM

10. Strunz, Bob, Waters, Mairead. 1999. The ELISE II Project, A Digital Image Library for Europe. Available online at URL: http://www.hut.fi/events/eunis99/Csession/C31.htm

Appendix A

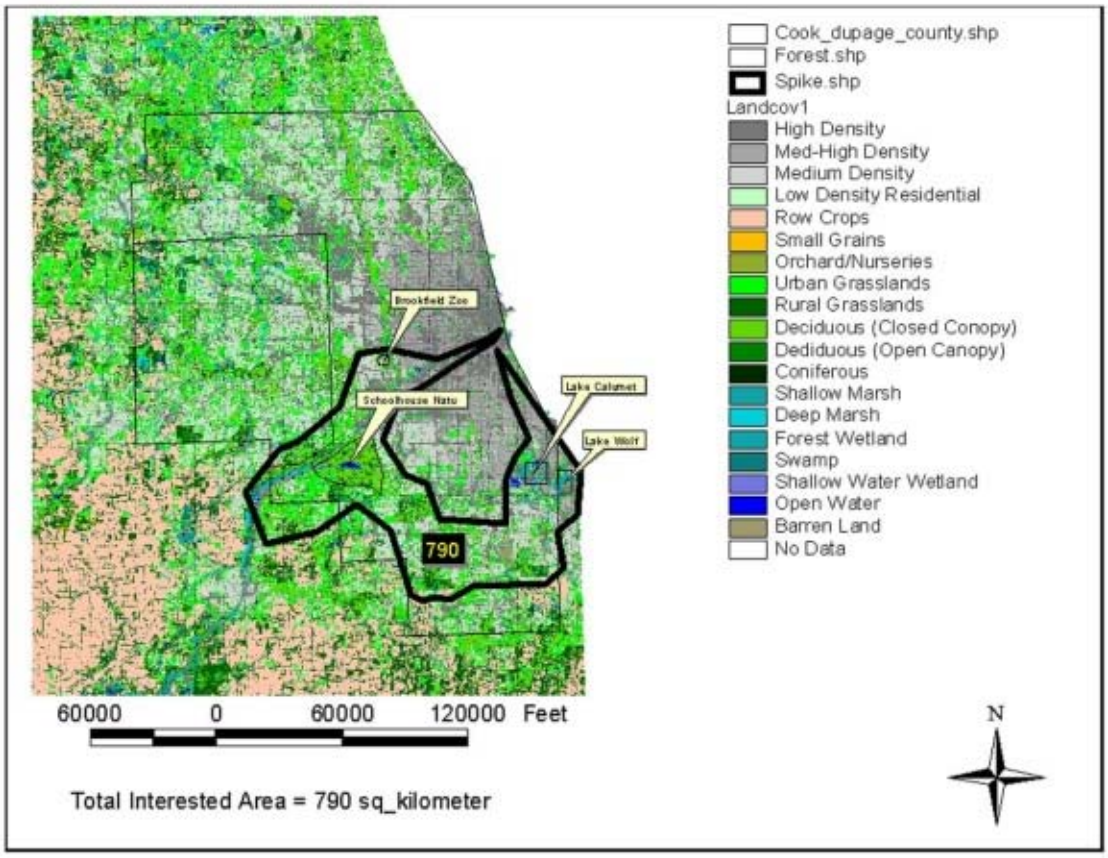

Figure 1: The study area in Chicago. The goal was to create a study area for student projects and future research on environmental issues in urban areas. The lines were gerrymandered to include several key sites. Interestingly, the center of this the location of a large cluster of cases of West Nile virus in 2002 and St. Louis Encephalitis in 1975.

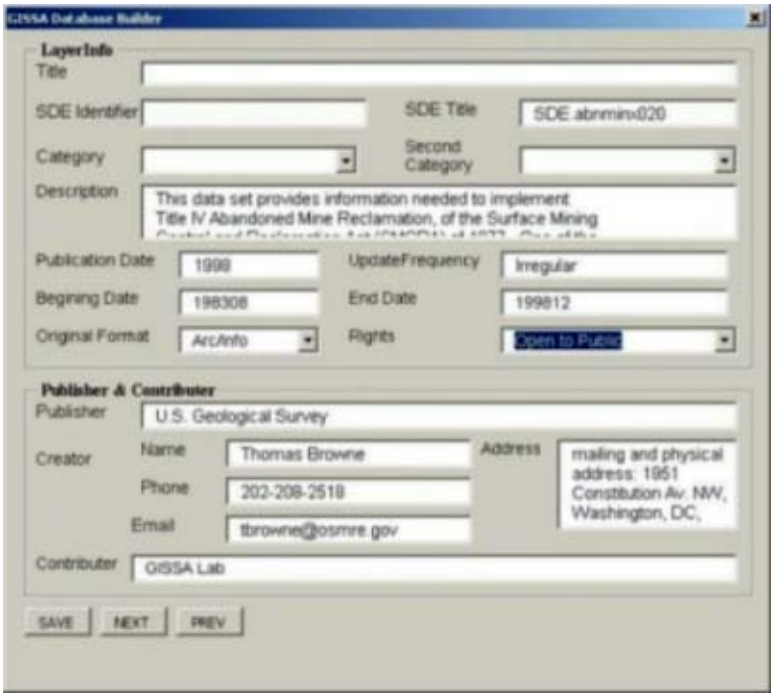

Fig 2: The Database Builder. From this screen, the administrator of the system enters in the revised Dublin Core Metadata fields for each layer. The information entered becomes the basis for the searches performed on the data in the system.

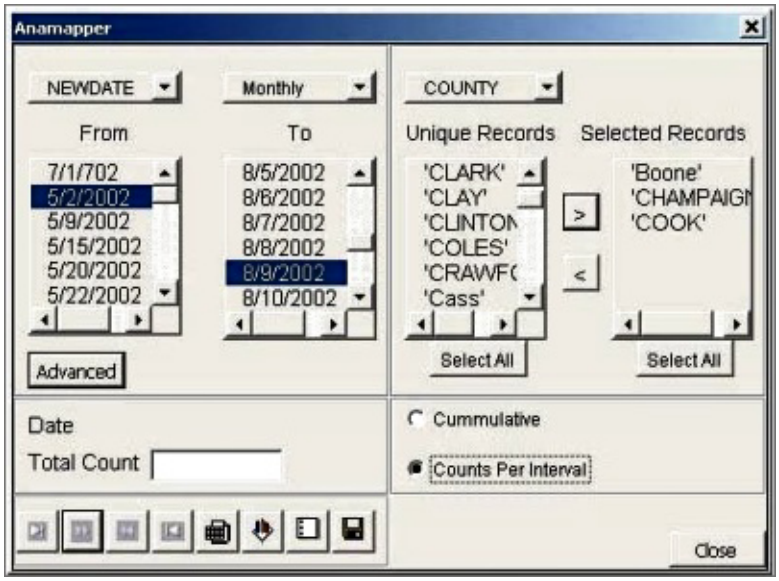

Fig 3: The AniMapper: The AniMapper is a temporal analysis tool developed GISSA Lab, Veterinary Medicine, UIUC.

Category	Secondary Category
Administrative	Political
	Census
	Public Lands
	Mapping
Natural Features	Hydrology
	Geology
	Elevation
	Soils
	Mines / Minerals
	Land Use / Land Cover
	Species Range & Habitat
	Hazards
Human Related	Transportation
	Utilities
	Agriculture inputs
	Crime
Demographic	Human
	Animal
Images	
Research Results	
Miscellaneous	

Table 1. Classification of spatial data into Primary and secondary category

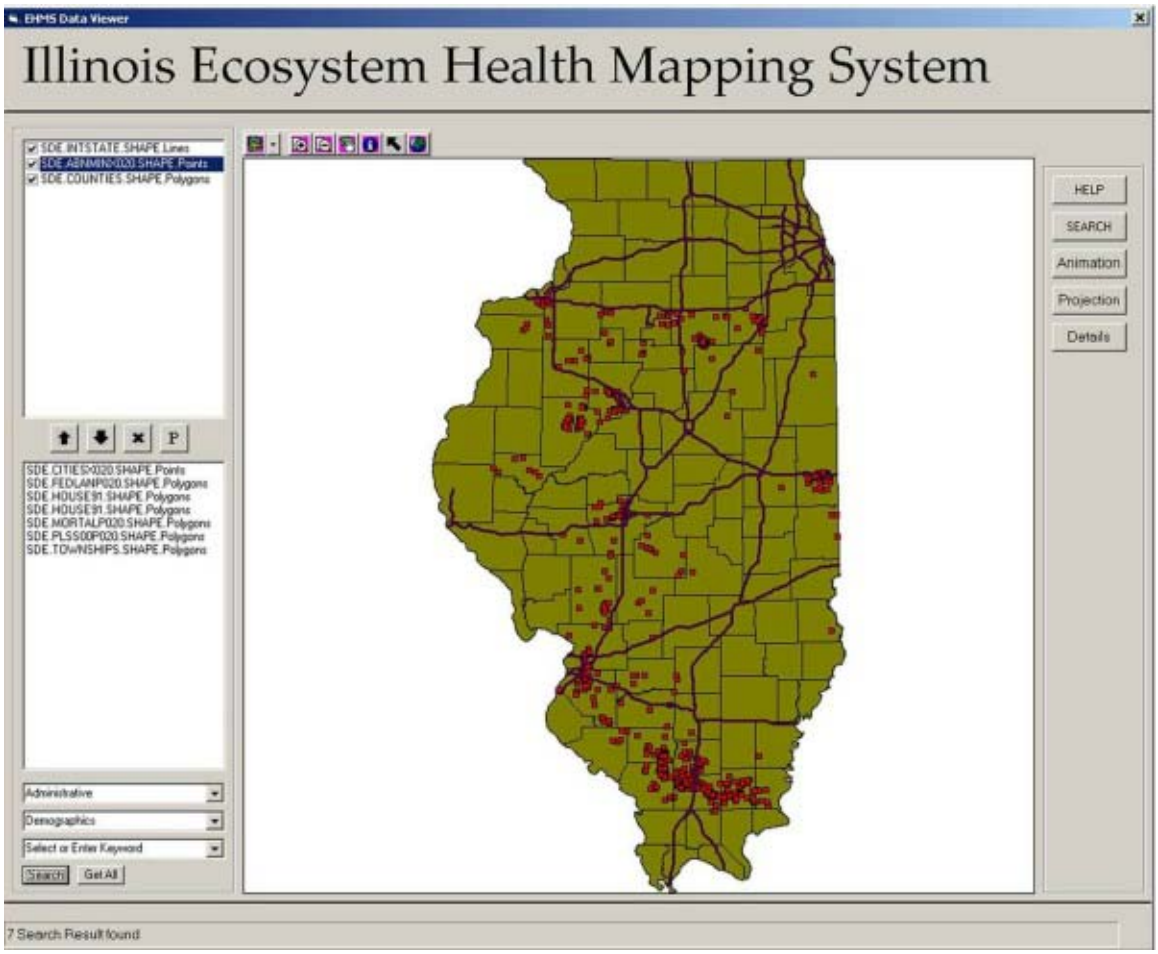

Figure 4: The EHMS Viewer: This is the viewer for SDE database. Beside general options like zoom in, zoom out, identity, it offers the facility to search SDE database and also do temporal analysis.